Multiple QTL mapping

Karl Broman

Biostatistics and Medical Informatics
University of Wisconsin – Madison

rqtl.org

kbroman.org

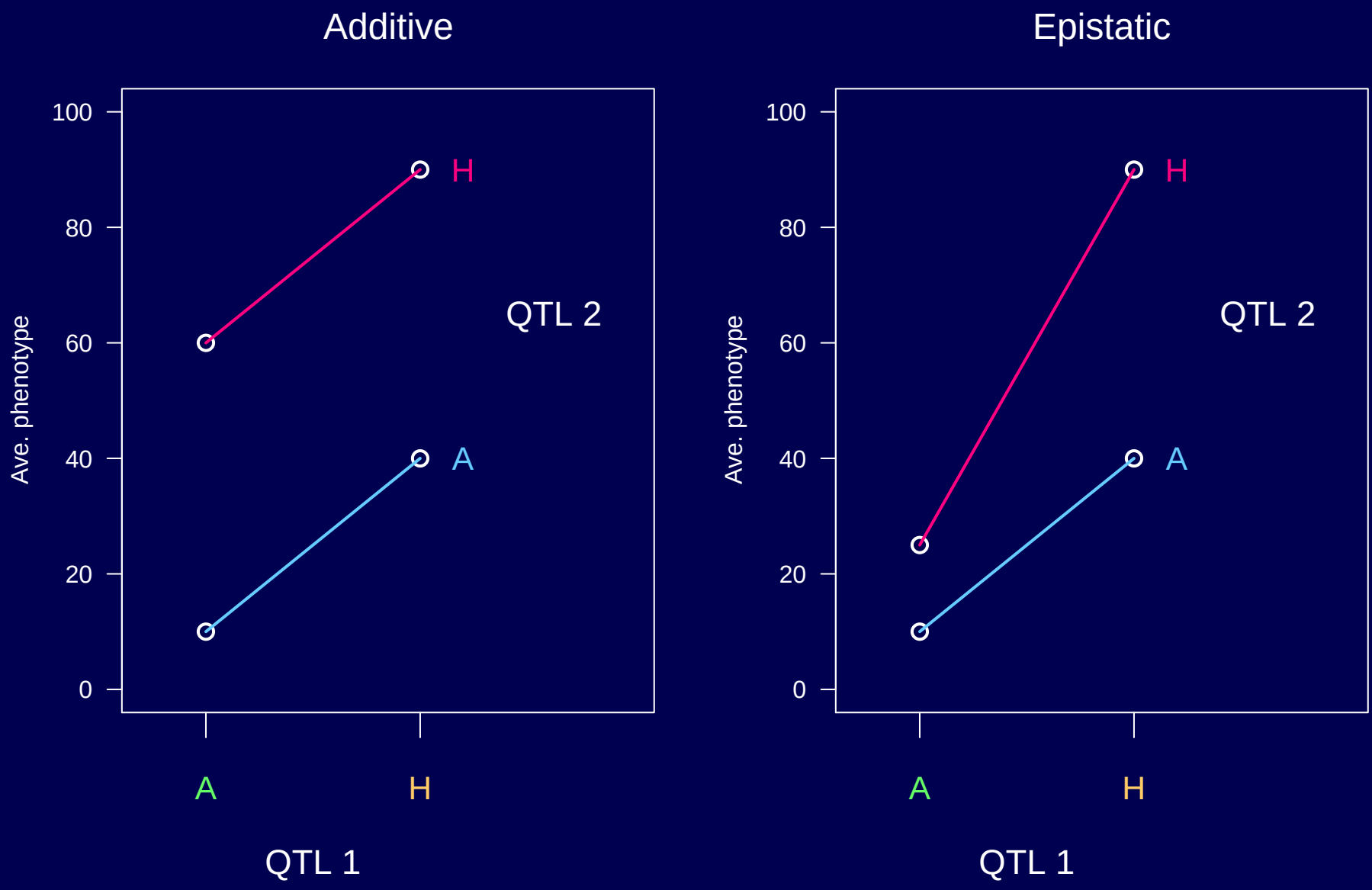
github.com/kbroman

@kwbroman

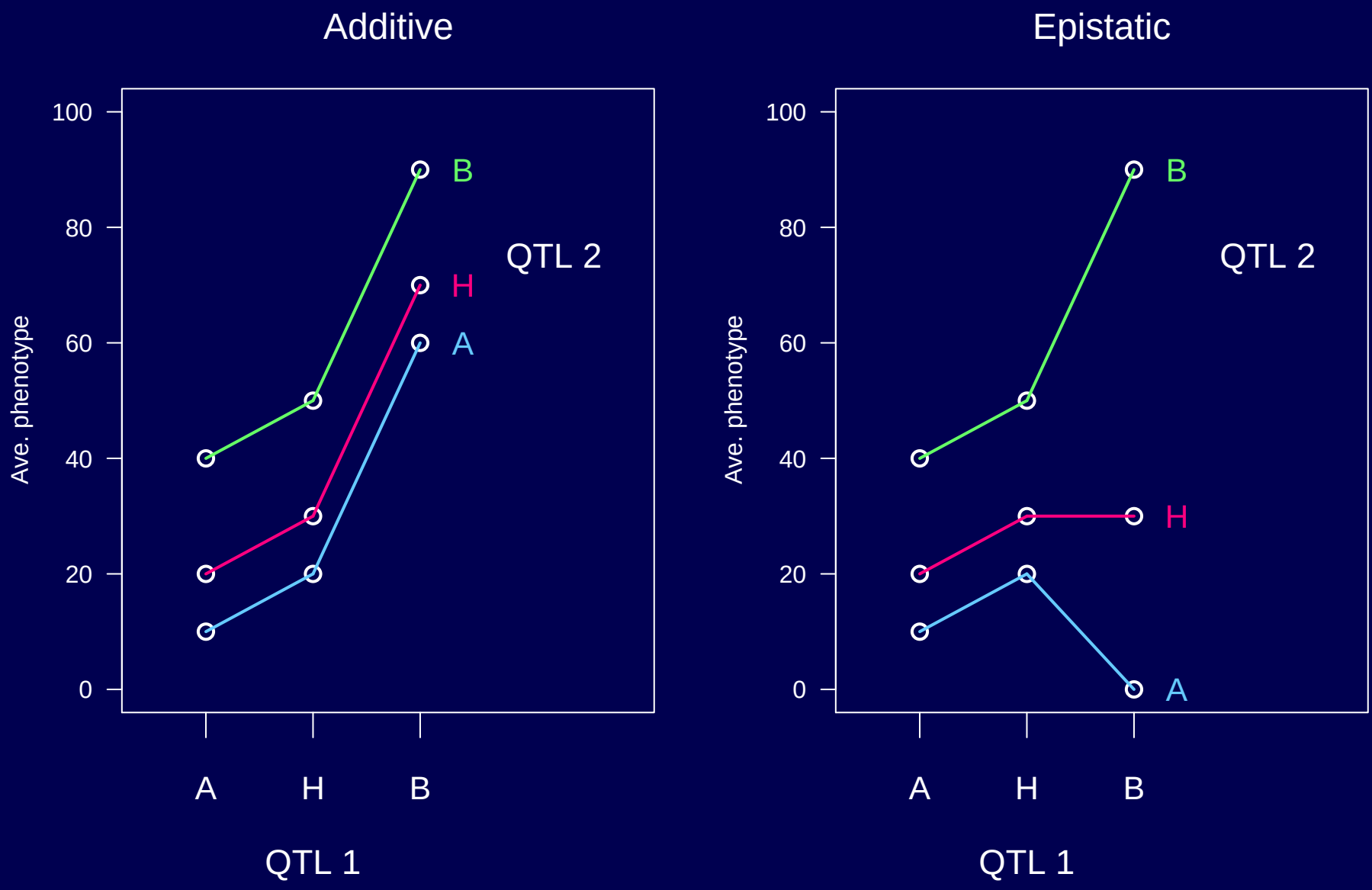
Modeling multiple QTL

- Reduce residual variation $\longrightarrow$ increased power
- Separate linked QTL
- Identify interactions among QTL (epistasis)

Epistasis in BC



Epistasis in F_2



2-dim, 2-QTL scan

For all pairs of positions, fit the following models:

$$H_f : y = \mu + \beta_1 q_1 + \beta_2 q_2 + \gamma q_1 q_2 + \epsilon$$

$$H_a : y = \mu + \beta_1 q_1 + \beta_2 q_2 + +\epsilon$$

$$H_1 : y = \mu + \beta_1 q_1 + \epsilon$$

$$H_0 : y = \mu + \epsilon$$

$\log_{10}$ likelihoods:

$$l_f(s,t) \qquad l_a(s,t) \qquad l_1(s) \qquad l_0$$

2-dim, 2-QTL scan

LOD scores:

$$\mathrm{LOD}_f(s,t) = l_f(s,t) - l_0$$

$$\mathrm{LOD}_a(s,t) = l_a(s,t) - l_0$$

$$\mathrm{LOD}_i(s,t) = l_f(s,t) - l_a(s,t)$$

$$\mathrm{LOD}_1(s) = l_1(s) - l_0$$

Summaries

Consider each pair of chromosomes, (j, k),
and let $c(s)$ denote the chromosome for position s.

$$M_f(j,k) = \max_{c(s)=j, c(t)=k} LOD_f(s,t)$$

$$M_a(j,k) = \max_{c(s)=j, c(t)=k} LOD_a(s,t)$$

$$M_1(j,k) = \max_{c(s)=j \text{ or } k} LOD_1(s)$$

$$M_i(j,k) = M_f(j,k) - M_a(j,k)$$

$$M_{fv1}(j,k) = M_f(j,k) - M_1(j,k)$$

$$M_{av1}(j,k) = M_a(j,k) - M_1(j,k)$$

→ R

- `scantwo()`
- `iplotScantwo()` in R/qtlcharts

Hypothesis testing?

- In the past, QTL mapping has been regarded as a task of hypothesis testing.

  Is this a QTL?

  Much of the focus has been on adjusting for test multiplicity.

- It is better to view the problem as one of model selection.

  What set of QTL are well supported?

  Is there evidence for QTL-QTL interactions?

  Model = a defined set of QTL and QTL-QTL interactions
  (and possibly covariates and QTL-covariate interactions).

Model selection

- Class of models
  - Additive models
  - + pairwise interactions
  - + higher-order interactions
  - Regression trees
- Model fit
  - Maximum likelihood
  - Haley-Knott regression
  - extended Haley-Knott
  - Multiple imputation
  - MCMC
- Model comparison
  - Estimated prediction error
  - AIC, BIC, penalized likelihood
  - Bayes
- Model search
  - Forward selection
  - Backward elimination
  - Stepwise selection
  - Randomized algorithms

Target

- Selection of a model includes two types of errors:
  - Miss important terms (QTLs or interactions)
  - Include extraneous terms
- Unlike in hypothesis testing, we can make both errors at the same time.
- Identify as many correct terms as possible, while controlling the rate of inclusion of extraneous terms.

What is special here?

- Goal: identify the major players
- A continuum of ordinal-valued covariates (the genetic loci)
- Association among the covariates
  - Loci on different chromosomes are independent
  - Along chromosome, a very simple (and known) correlation structure

Exploratory methods

- Condition on a large-effect QTL
  - Reduce residual variation
  - Conditional LOD score:

$$\text{LOD}(q_2 \mid q_1) = \log_{10}\left\{\frac{\Pr(\text{data} \mid q_1, q_2)}{\Pr(\text{data} \mid q_1)}\right\}$$

- Piece together the putative QTL from the 1d and 2d scans
  - Omit loci that no longer look interesting (drop-one-at-a-time analysis)
  - Study potential interactions among the identified loci
  - Scan for additional loci (perhaps allowing interactions), conditional on these

→ R

- `scanone()` with marker as additive covariate
- `makeqtl()`, `fitqtl()`, `addqtl()`, `refineqtl()`

Automation

- Assistance to non-specialists
- Understanding performance
- Many phenotypes

Additive QTL

$$y = \mu + \sum \beta_j \, q_j + \epsilon \qquad \text{which } \beta_j \neq 0?$$

$$\text{pLOD}(\gamma) = \text{LOD}(\gamma) - T\,|\gamma|$$

Additive QTL

$$y = \mu + \sum \beta_j q_j + \epsilon \qquad \text{which } \beta_j \neq 0?$$

$$\text{pLOD}(\gamma) = \text{LOD}(\gamma) - T\,|\gamma|$$

0 vs 1 QTL: $\text{pLOD}(\emptyset) = 0$

$$\text{pLOD}(\{\lambda\}) = \text{LOD}(\lambda) - T$$

Additive QTL

$$y = \mu + \sum \beta_j q_j + \epsilon \qquad \text{which } \beta_j \neq 0?$$

$$\text{pLOD}(\gamma) = \text{LOD}(\gamma) - T\,|\gamma|$$

For the mouse genome:

T = 2.69 (BC) or 3.52 (F_2)

→ R

- `stepwiseqtl()`
- `plotLodProfile()`

References

- Strickberger MW (1985) *Genetics*, 3rd edition. Macmillan, New York, chapter 11.

  An old but excellent general genetics textbook with a very interesting discussion of epistasis.

- Broman KW, Speed TP (2002) A model selection approach for the identification of quantitative trait loci in experimental crosses. J Roy Stat Soc B 64:641–656

  Multiple-QTL model selection with additive QTL.

- Manichaikul A, Moon JY, Sen Ś, Yandell BS, Broman KW (2009) A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. Genetics 181:1077–1086

  Also account for epistasis.